The Asian Journal of Technology Management Vol. 15 No. 1 (2022): 51-68

Understanding Customer Insights Through Big Data: Innovations in Brand Evaluation in the Automotive Industry

Batara Parada Siahaan[1*] and Eko Agus Prasetio[2]
[1]Bachelor Of Applied Science in Software Engineering Technology Program,
Del Institute of Technology, Indonesia,
[2]School of Business and Management, Institut Teknologi Bandung, Indonesia

Abstract. *Insights gained from social media platforms are pivotal for businesses to understand their products' present position. While it is possible to use consulting services focusing on surveys about a product or brand, such methods may yield limited insights. By contrast, on social media, people frequently express their individual and unique feelings about products openly and informally. With this in mind, we aim to provide rigorous methodologies to enable businesses to gain significant insights on their brands and products in terms of representations on social media. This study employs conjoint analysis to lay the analytical groundwork for developing positive and negative sentiment frameworks to evaluate the brands of three prominent emerging automotive companies in Indonesia, anonymized as "HMI," "YMI," and "SMI." We conducted a survey with a sample size of n=67 to analyze the phrasings of importance for our wording dictionary construction. A series of data processing operations were carried out, including the collection, capture, formatting, cleansing, and transformation of data. Our study's findings indicate a distinct ranking of the most positively and negatively perceived companies among social media users. As a direct management-related implication, our proposed data analysis methods could assist the industry in applying the same rigor to evaluating companies' products and brands directly from social media users' perspective.*

Keywords: *Brand image, social media, data analytics, sentiment analysis, conjoint analysis*

1. Introduction

Developments in terms of information on social media indicate that communication channels are transitioning from an offline to an online state. Each day, millions of textual records are created as users communicate more indirectly through the channels which large Internet corporations such as Facebook and Twitter offer them. Individuals can use Internet platforms such as social networks to communicate their feelings about their daily lives and, more crucially for business studies, about activities relating to a company, service, product, or brand (Jian & Ping, 2015). These insights are invaluable since the information is real-time, highly personal, and reflects the true position of a company's product, service, and image in the eyes of the consumer, often serving to represent the client's wishes for the product or service at hand (Kencanasari et al., 2021; Jian & Ping, 2015; Wang & Wang, 2015).

Free, accessible, and available insights are essential for businesses' decision-making and market analytical purposes (Kumar & Kaur, 2016; Wang & Wang, 2015). While various services aiming to help companies obtain customer insights are available, such as the use of a marketing consultant to conduct surveys or conversations with customers, the breadth of perspectives they offer statistically pales in comparison to the massive volume of written opinions posted on social media. In other words, gaining insights from social media can be seen as a tactical solution due to the abundant directional information present there (Hemalatha et al., 2013; Wang & Wang, 2015).

*Corresponding author. Email: batara_parada@sbm-itb.ac.id
Received: December 16th, 2021; Revised: April 22th, 2022; Accepted: April 22th, 2022
Doi: http://dx.doi.org/10.12695/ajtm.2022.15.1.4
Print ISSN: 1978-6956; Online ISSN: 2089-791X.

Rather than waiting for onsite consumer complaints, it is preferable for a competitive corporation to actively seek market insights. Actionable decisions are facilitated by the massive volume of directional information available on the Internet in the form of feelings and user reviews (Hemalatha et al., 2013). With millions of words, friendly or nasty remarks, favorable or unfavorable posts, and comments with high interaction from other users being regularly posted, including widespread, "viral" social media content, a company's image can be rapidly and severely impacted. The visible factor in social media data is customer interaction; as users are more likely to reveal their true selves on social media than under other circumstances, for knowledge management, businesses ought to leverage this data to achieve innovation (Helmi, 2020)

Users of social media platforms such as Facebook and Twitter can share the same post across multiple circles of the network, which is freely accessible online. Companies can employ opinion mining toward actively monitoring the consumer experience, as it is an essential aspect for business (Dong et al., 2015). Product positioning can be analyzed through big data, which can prove vital in a competitive market alongside the necessity to incorporate data mining into the decision-making process (Kumar & Kaur, 2016). Furthermore, the position of this paper is essential because there are several critical aspects related to the approach and object under consideration. Our literature review reveals a research gap in the judgment analysis approach using a dictionary as its basic construction with statistical validation. Although some data-mining studies have employed dictionary-based tokenization (see e.g., Khder & Yaseen, 2017; Reagan et al., 2017), we choose to provide an alternative for managing large datasets by applying a part-worth-utility and conjoint analysis. Many researchers have been interested in the analysis of negative and positive validation of sentiment in determining the position of a word, and conjoint analysis can be used as the utility basis.

In this research, we provide a dictionary-based foundation using conjoint analysis. This means that the basis of negative and positive words becomes measurable in terms of utility. The appearance of these words on social media will serve to indicate the negative position of a review of the brands under examination, namely three Indonesian automotive companies. While data-mining studies generally can explain whether a sentiment is negative or positive, this study aims to not only describe the position of a sentiment between negative and positive, but also to show the positive and negative levels of a sentiment.

In summary, our attempt here is to propose an understanding of how to retrieve data from social media, which could facilitate companies' assessment of the present market from customers' perspective. Thus, we present our research questions as follows: Research Question (RQ) 1: How does one examine the product, service, or brand by using sentiment analysis through big data?

Research Question (RQ) 2: How does sentiment analysis help companies evaluate the market from customers' perspective toward their competitors' position?

The following sections lay out this study's contents: After having established the research framework, theory, context, aims, and research questions in the "Introduction," we proceed to the "Literature Review," listing preparatory studies pertinent to the present context. The "Methodology" then addresses the methods we used to conduct this study, followed by the "Findings and Discussion" section detailing our results and approaches, such as statistical and trend analyses, statistics generated from evaluation, and data analytics. The "Conclusion" offers final considerations.

2. Literature Review

Brand evaluation is a pivotal mechanism in performance management. However, the route to its implementation is sometimes fraught with difficulties, due to its being based on intangible consumer perceptions in terms of choices. Hence, amid its literature review, this study endeavors to assess the scale of brand evaluation and customer insight implementation in relation to the application of big data analytics, which corresponds to the present methodology.

Big data applications in brand evaluation have mostly been related to the text-mining process with a case-based structure, such as in Youn et al. (2020), who drew from big data analytics to evaluate Hyundai Motor's brand. The researchers employed data mining with morphological analysis and semiotics to process data into information with high utility, with the purpose of understanding the data's sensibilities toward boosting the brand element (Youn et al., 2020).

An effective brand evaluation depends on the speed and ease of data access, which makes big-data technology particularly useful when compared to traditional methods such as the application of PEI to gauging brand image (Saura et al., 2019). According to Saura et al. (2019), while such traditional applications might have remained relevant and usable, the vast scalability of information on social media could offer a faster route to obtaining a much larger breadth of perceptions.

Relevant and widely available data sources are critical to evaluating brands and obtaining customer insights, as demonstrated by our use of Facebook as a data source. Other researchers have drawn from different social networks, such as Shirdastian et al. (2019), whose case study relied on Twitter data to assess the Starbucks brand; and Lee et al. (2015), who employed image-based social media platforms (e.g., Instagram) to understand brands and customer insights.

The incorporation of big data analytics has often been related to brand evaluation and customer insights. Mariani and Wamba (2020) applied it in measuring a company's innovativeness; Shirdastian et al. (2019) in the context of brand authentication; and Chiang and Yang (2018) in evaluating country origin in relation to personality brands. The wide availability of data on social networks has motivated various researchers to seek out information online for the management and evaluation of brands and customer insights, ranging from strategic brand management (Tirunillai & Tellis, 2014) to the prediction of insights for customer demand (Chong et al., 2017). Such literature speaks to the relevance of data source availability, analytical context, and research purpose for brand evaluation.

From the analytical perspective, we ought to discuss sentiment analysis as another tool for brand evaluation. Sentiment analysis can rely on a variety of methods, including machine learning and lexicon-based analysis, or "rule-based analysis" (Medhat et al., 2014). Various difficulties have been associated with the application of sentiment analysis, however, especially in terms of achieving accuracy and of the limitations of available approaches, including machine learning and lexicon- or rule-based methods (Hussein, 2018). As the major issues affecting sentiment analysis research, Shayaan et al. (2018) identified the unstructured, heterogenous nature of online data and cases of data volume insufficiency necessitating transformation.

Hence, difficulties in the approach have included the availability of analytical material and data volume, besides low levels of Internet penetration and filtering—which have tended to limit the relevant data sources—and multilingualism and language-related issues in processing (Smetani, 2020). The way language is treated affects the type of algorithm that can be used to analyze the thoughts collected. In summary, the main obstacles in sentiment analysis have been obtaining accuracy, ensuring data availability, and proposing diverse methodologies which can accommodate a variety of problems associated with big data aspects (Alsayat & Elmitwally, 2020; Alwakid et al., 2017).

It can be said that natural language is extremely casual in comparison to human brain operations, yet extremely limiting in comparison to human-created algorithms (Dong et al., 2015). As a result, each feature considered to aid in the comprehension of big data analytics requires substantial explication. For the automotive industry, which includes automobiles and motorbikes, transportation choices tend to be based on individual rather than shared use, making public opinion essential. This perspective is referred to as the "brand image" or "product image" in the present study.

As Shayaan et al. (2018) proposed, amid the purposes of sentiment analysis, several issues have tended to arise in terms of data format, linked to a failure to develop and gather data that is ready for analysis. Our literature review demonstrates that the research gap and obstacles associated with sentiment analysis remain substantial. Additionally, the gap indicates that, in terms of accuracy, we should pay attention to the contribution of methodologies that can quantify accuracy.

For our data processing, we construct a dictionary based on positive and negative words in the Indonesian language. Like various other studies, the present one adopts a hybrid strategy, applying statistical tools to determine not only word composition, but also sentiment levels. We attempt to process data in a traceable manner through the lexicon-based method. The relevant rule-based methods are expected to be applicable to any sector as long as the data items required for the research are available.

Many believe that grasping social media information is critical to grasping real-world circumstances, whether in business, society, or politics (Akter & Azis, 2016; Sandoval-Almazan & Valle-Cruz, 2018; Soliman et al., 2014; Troussas et al., 2013). In employing particular analyses in gathering and creating understanding about social media via a robust algorithm, a competitive advantage can be gained. The importance of data such as feedback on social media is that it tends to depict users' feelings related to a product or service, as users usually provide remarks when they encounter a problem or when the product or service exceeds their expectations (Kencanasari et al., 2021).

In another vein, Khedr and Yaseen (2017) attempted to forecast the stock market by mining and analyzing data from news packages and converting them to sentiment, through news tokenization and a naive Bayes algorithm. The process of comprehending large-scale text has been known to require a great deal of resources, while the tokenized context of a sentence could suffice (Reagan et al., 2017). Tokenization can be used to aid in the comprehension of sentence structure by categorizing words as positive or negative (Reagan et al., 2017). The tokenization process is crucial when quantifying the frequency of multiple positive sentiments and pessimistic views is required. Apart from the concept of natural language, which allows for mistyping and errors, and conceptual machine learning, which enables them to tolerate mutual understanding (Khoo et al., 2012; Kralj et al., 2015; Wang & Wang, 2015).

Our study emphasizes the structure for constructing a wording dictionary. Its value is validated via conjoint analysis, which in turn emphasizes the importance of conducting a survey to obtain user feedback on a variable tested in the form of a ranking. A major implication is that conjoint analysis can be used in a variety of disciplines, particularly those involving the testing of customer perceptions regarding certain business.

According to Hauser and Rao (2004), as a multidimensional scale measuring business marketing perceptions, conjoint analysis has gained importance across business domains. Various conceptual and empirical studies related to marketing and product development, such as the product-positioning study by Green and Krieger (1993), and various studies related to perception metrics on services in the health sector, as reported in the study by Marshall et al (2010).

Table 1.
State of the Art

Author (Year)	Data Mining	Token	Mathematic Modeling	Dictionary Based	Statistics Wording	Judgment Analysis
Khedr & Yaseen (2017)	✓	✓	✓			
Reagan et al. (2017)	✓	✓		✓		✓
Dong et al. (2016)	✓		✓			✓
Jian & Ping (2015)	✓	✓				
Hemalatha et al. (2013)	✓			✓		✓
Mejova (2012)	✓					
This Study	✓	✓		✓	✓	✓

In a comprehensive study, Cattin and Wittink (1982) posited that conjoint analysis had diversified into business development, such as the areas of product development, market segmentation, pricing, and distribution; as well as having spread to the consumer goods, government, services, and transportation sectors. More recent research has found conjoint analysis to have developed in the direction of measuring the perception of acceptance of a product or service, due to its ability to gauge attributes multidimensionally (Papadima et al, 2020). For example, amid the COVID-19 pandemic and emphasis on vaccination implementation as a goal for many countries, Sun et al. (2020) used conjoint analysis to measure vaccination preferences by region—which corroborates the present trend of employing conjoint analysis in gauging preferences.

On the topic of surveys, Cattin and Wittink (1982) categorized the method of applying conjoint analysis to survey collection into full profile, two factors, and profile combination. Due to this multidimensionality, researchers can more easily measure various attributes, by combining the attributes being measured. For the present study, we measure wording based on a combination of profiles, since social media users can express two or more positive sentences in a single comment on platforms such as Facebook.

We find conjoint analysis to be one of the most common references across empirical studies on brand perception and customer insights. Conjoint analysis has been applied in addressing customer preferences (Tripathi and Siddiqui, 2010), customer satisfaction (Lim et al., 2021), branding (Eggers et al., 2016), among many other factors of brand evaluation and customer insight development (Youn et al., 2020; Saura et al., 2019; Chiang & Yang, 2018). This indicates that conjoint analysis is highly pertinent to the purposes of the present study, in which we evaluate company brands in Indonesia's automotive industry.

As shown in Table 1, this study's innovative contribution builds on previous research, particularly in terms of sentiment analysis. Our literature review indicates that data-mining studies have devoted much attention to tokenization, and that other studies have used a dictionary in a similar manner to ours. However, most studies evaluating sentiment did not test validation for the scale of how the dictionary was statistically measured. By combining data-mining goals with judgment analysis models, conjoint analysis can enable strong validation for the dictionary construct created from the big data drawn upon, serving as our main novel contribution.

It is also worth noting that the variety of data formats and inconsistencies make it considerably difficult to define and process all the types of data in an algorithm; for example, processing photos is quite different from processing text (Kralj et al., 2015). Hence, we decided to ensure that this study would be sufficiently rigorous by limiting it to text-based data mining. Other sorts of data, such as photographs, emoticons, and videos, are therefore omitted from the study.

The more advanced the technology, resources, and algorithms are, the more efficient the analytics tend to be, although different sorts of data cannot usually be combined in the same method (Lin et al., 2017; Wang & Wang, 2015). For instance, in image analytics, the preprocessing of database semantics is not applicable to text- or video-based data (Lin et al., 2017).

3. Methodology

Figure 1 summarizes our research process, which includes (a) literature evaluation, (b) conjoint analysis, (c) data processing, (d) benchmarking analysis, and (e) conclusion.

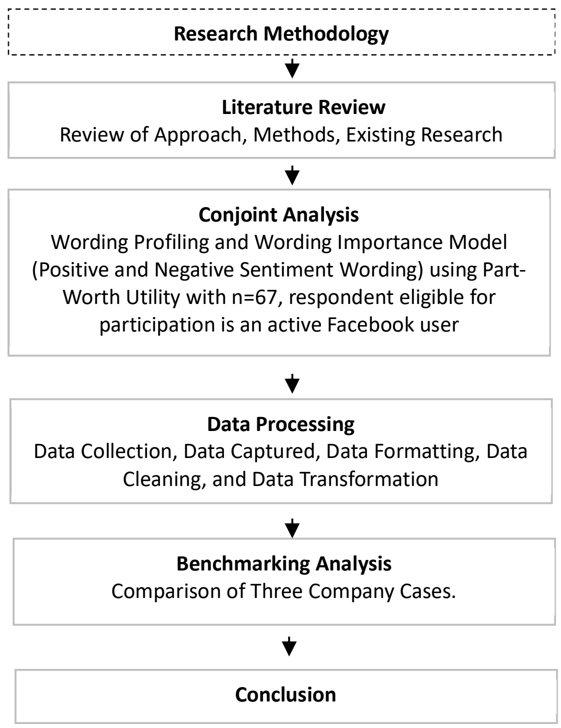

Figure. 1.
Research Methodology

Literature Review
In the "Literature Review" section, we examine the body of knowledge derived from data-mining studies, particularly the evaluation of digital content via big data. It addresses the general definition of big data, gap analysis in context, as well as methods, followed by a review regarding aspects of our study's contribution, such as the usefulness of conjoint analysis in applying big data to gauge companies' products or brands.

Conjoint Analysis
We propose statistical analysis to assess the value of each language framework design, while developing a framework for positive and negative emotion. We have selected conjoint analysis as our approach because it can eliminate the ambiguity associated with subjects' psychological elements as a means to determine preferences for a measurable object (Bak & Bartlomowicz, 2009).

Table 2.
Positive and Negative Wording (Z-Axis)

Positive Wordings Samples	**Negative Wording Samples**
(code 1) Love this product	(code 1) Hate this product
(code 2) This is a nice product	(code 2) This stupid product
(code 3) Cool product	(code 3) Bad product
(code 4) Wonderful product	(code 1) I don't like this product
(code 1) Like this product	(code 2) This is horrible
(code 2) Amazing product	(code 3) The use of negative jargon
(code 3) Great product	
(code 4) Best product	

Conjoint analysis' data processing in R entails three distinct components of data-form design as prerequisites for conducting the study: the X-axis, Y-axis, and Z-axis. The X-axis represents profile combination, the Y-axis depicts respondents' preferences, and the Z-axis depicts the qualities measured. To apply conjoint analysis in this study, we must first establish the Z-axis, which pertains to the wording dictionary classifying sentiment (Table 2). The Z-axis explains positive and negative wording, based on the study's lexicon. For this analysis, we classify eight positive and six negative word expressions.

A remark being analyzed can vary from the "I love this product" (code 1) represented in Table 2, for instance; it may also be "I love their service," or "their company is lovely," as long as the sentence involves the word "love" in a positive polarity. We must also consider that negative sentiment may not necessarily be expressed as "I hate this product." However, while the wording may differ, "hate" retains its negative polarity. Given the importance of accurately measuring the aims of these positive and negative categories,

Upon determining the positive and negative wording (Z-axis), the next step is to combine the wording (X-axis), as Table 3 shows.

P1 is the abbreviation of "Profile 1" in Table 3. Profile code can be determined by dividing the positive wording samples in Table 1. In P1, for instance, the combination code is 1,1, which pertains to the combination between "Love this product" (code 1) and "Like this product" (code 1). Thus, when these sentences are combined, there will be P1-P16 (16 combinations) for positive words and P1-P9 (9 combinations) for negative words.

Table 3.
Wording Combination (X-axis)

Positive Wording Combination				**Negative Wording Combination**			
P1	1,1	P9	3,1	P1	1,1		
P2	1,2	P10	3,2	P2	1,2		
P3	1,3	P11	3,3	P3	1,3	P7	3,1
P4	1,4	P12	3,4			P8	3,2
P5	2,1	P13	4,1	P4	2,1	P9	3,3
P6	2,2	P14	4,2	P5	2,2		
P7	2,3	P15	4,3	P6	2,3		
P8	2,4	P16	4,4				

After obtaining the word combination (X-axis) displayed in Table 3, we calculate the part-worth utility of the combination to determine the most positive and negative word combinations. A survey can be used to accomplish this.

We work with 67 original samples acquired via offline (29 samples) and online surveys (38 samples); because no data is missing, the data for processing equals 67 samples, which is sufficient for conjoint analysis (Green & Krieger, 2002). Table 4 summarizes the samples' demographics. As the criteria for participation in this study, respondents must be familiar with the Facebook social network, active as users, and at least 18 years old.

Table 4.
Demographics of Respondents

Gender	Online Survey	Offline Survey	Total
Female	20	20	40
Male	18	9	27
Age Range (20–30)	24	29	53
Age Range (31–50)	14	0	14
Total	38	29	67

Respondents were given the combination profiles in Table 3 (X-axis) and asked to rate them on a scale of 0 to 10. The survey briefly explained that this language would be used in a scenario in which users expressed their personal feelings about a company's product, service, or image in their social media comments, whether favorably or unfavorably. A score of 0 indicates that the profile combination is not strong, while 10 indicates that the statement highlights the strongest emotion or most significant expression, whether of a positive or negative attitude.

Data Processing

Data processing entails the steps of data collection, capture, formatting, cleansing, and transformation (Figure 2).

The first step of data processing is data collection, in our case using the PHP Scripting Programming V.7.2.4 from the XAMPP open-source package; and collecting data from the official Facebook pages of three automotive companies, which we will call "YMI," "SMI," and "HMI." To avoid future conflicts over the sensitive market analysis in this study, we choose to anonymize the firms' names by assigning them pseudonyms. We analyze users' comments on each company's Facebook page, using Facebook's GRAPH API (with certain data limitations), having gathered around 10,000 comments from each page.

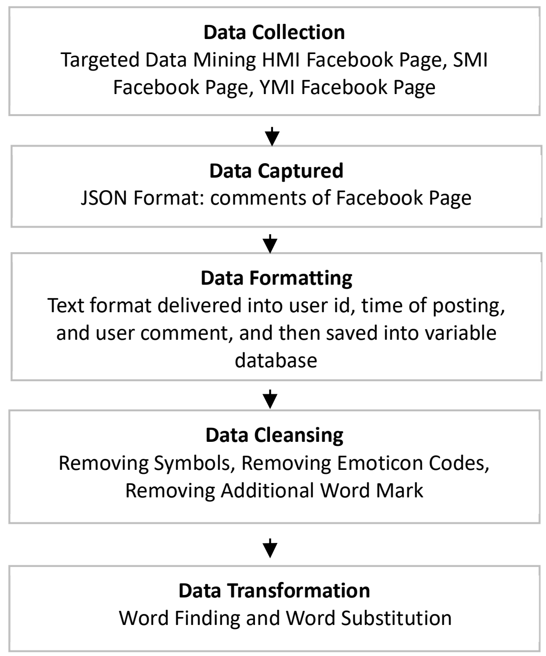

Figure 2.
Data Processing Steps

The second step of data processing is data capture, done by saving it to a database. The data captured was still unreadable and in the JSON format, needing further processing.

The third step is data formatting, in which all data to be read is put into a readable format. JSON must be formatted into "user id," "time of posting," and "user comment," and then saved into a variable database.

The fourth step is data cleansing, whose purpose is to eliminate superfluous emoticon code, symbols, and incomplete, meaningless words (Dong et al., 2015). We first analyze the structure of the raw data acquired, using naive judgment to examine the words, sentences, and how the term was expressed holistically. This judgment is then employed to identify which words in the sentences are unclear, incomplete, or should be deleted (Wang & Wang, 2015). In developing an algorithm for the data-cleansing process, we could construct a feature in the algorithm allowing us to group each word distinctively and holistically, and to present each word as a statistical number.

The last step of data processing is data transformation, converting incomplete words to words based on an user's individual input. Without transformation, data will be incomplete; certain abbreviations may skew the sentiment's polarity dependent on the search feature's frequency. As a result, all abbreviations and strange terms must be eliminated or replaced with a more "comprehensible" term. This stage will review and replace missing, incorrectly spelled, and unclear comments with their correct counterparts.

Benchmarking Analysis
Benchmarking analysis is one of the main objectives of this study; through it, the company's positioning can be obtained via comparison to a competitor. The goal is to determine which companies have been the most positively represented (going from the most positive to the least positive).

From our understanding, the context of positive and negative sentiment is comparable across the companies, as the competitive market characteristics are the same. We use the utility rate total calculated from the results constructed from the conjoint analysis, rated according to the frequency of positive and negative results. We assume that the majority of the frequencies associated with the appearance of positive and negative sentiment could be measured statistically and that benefit could eventually be derived from these steps in market competitors' decision-making via data mining (Kumar & Kaur, 2016).

4. Findings and Discussion

This chapter first discusses the results from the conjoint analysis, as the basis for the positive and negative terminology. The second aspect being discussed is the outcome from big data analytics, and the third is that of the benchmarking analysis (mentioned in the "Methodology" section).

Importance Model of Polarity
The outcome from positive phrasing is presented via the conjoint analysis results' worth utility. The statistical study performed through the R software package entailed crucial measurement of each positive wording, with the following fit model results for positive sentiment: 1,456e-10 p-value, R-square 0,0525, RSE 2,557 on 1064 DF, Median 0,1167.

Table 5.
Summary of Model Results for Positive Sentiment

Coefficient	**Estimate**	**Std. Error**	**t value**	**Pr (>|t|)**
Intercept	5,25105	0,07814	67,200	< 2e-16 ***
Positive Sentiment Group 1.1	0,89074	0,13530	6,583	7,21e-11 ***
Positive Sentiment Group 1.2	-0,04956	0,13530	-0,366	0,7142
Positive Sentiment Group 1.3	-0,28463	0,13530	-2,104	0,0356 *
Positive Sentiment Group 2.1	0,42852	0,13547	3,163	0,0016 **
Positive Sentiment Group 2.2	-0,08314	0,13530	-0,614	0,5390
Positive Sentiment Group 2.3	-0,14284	0,13530	-1,056	0,2913

We can view the significance of those terms from the lenses of the findings from user psychology surveys, using utility-worth statistical models. Results indicate that "love" is the most powerful expression of positive feeling, followed by "great," "amazing," and "wonderful" (part-worth: love = great + amazing + wonderful).

Table 5 displays the consistent outcome for utility worth. For Group 1.1 in positive sentiment, the Pr value is significant at the *** level. For Group 2, the strongest expression is "like," followed by "nice," "cool," and "best" (part-worth: like = nice + cool + best). At the ** level, the Pr value is significant (Group 2.1).

In Figure 3, the sample bar depicts the psychological ranking of the importance of positive emotion based on the survey data. According to the bar, use of the word "love" is preferred over that of "great," "amazing," and "wonderful." The implication in Figure 3 is as follows: When two people both express they like a product and the first person uses the word "love" and the second person uses the word "great," the conjoint analysis indicates that the first person is being more positive than the second.

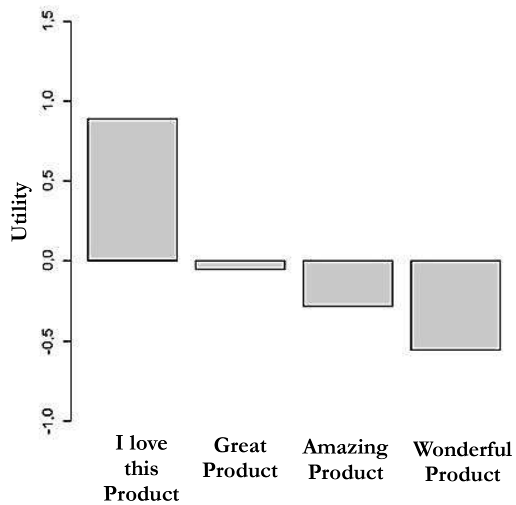

Figure. 3.
Utility Diagram for Positive Polarity Sentiment

This framework can minimize complexity in linguistic preferences. Evaluating the breadth of sentiment from a large number of phrases containing several expressions tends to be impossible without consuming a large volume of resources for a sophisticated algorithm, and this solution offers a method to analyze the big data saved.

Table 6.
Summary of Model Results for Negative Sentiment

Coefficient	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Intercept	4,93532	0,09025	54,685	< 2e-16 ***
Negative Sentiment Group 1.1	1,16915	0,12763	9,160	< 2e-16 ***
Negative Sentiment Group 1.2	-0,42289	0,12763	-3,313	0,000978 ***
Negative Sentiment Group 2.1	0,42786	0,12763	3,352	0,000852 ***
Negative Sentiment Group 2.2	-0,11940	0,12763	-0,936	0,349898

We present the results for negative wording based on the statistical analysis done via the R software package. This pertains to critical measurement of each negative wording with the fit model results for negative sentiment: 1.456e-10 p-value, R-square 0.1408, RSE 2,216 on 598 DF, Median, 0.0149

According to this utility-worth statistical model, the strongest expression of negative sentiment is "hate," followed by "bad" and "don't like" (part-worth: hate = bad + don't like). The consistent outcome indicates that the Pr value is significant at the *** level for Group 1.1 in negative sentiment (Table 6). "Stupid" is the strongest expression of negative sentiment in Group 2, followed by "horrible" and "negative jargon" (part-worth: stupid = horrible + negative jargon). The consistent outcome indicates that the Pr value is significant at the *** level for Group 2.1 in negative sentiment (also in Table 6).

In Figure. 4, we rank the psychological component to facilitate the use of conjoint analysis. The figure demonstrates that in expressing a negative attitude regarding a company's product, image, or brand, the term "hate" is favored over "bad" and "don't like," and "stupid" is preferred over "awful" or "jargon" or "sexual object."

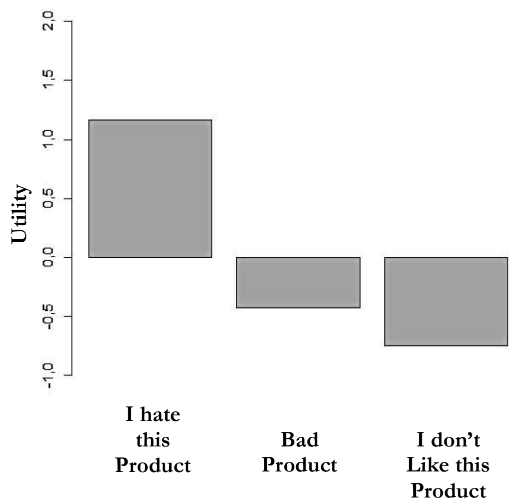

Figure. 4.
Utility Diagram for Negative Polarity Sentiment

Data Analytics
Here we provide the data-mining findings from big data analytics on the Facebook pages of the three Indonesian motor vehicle companies "HMI," "SMI," and "YMI." Regarding positive sentiment toward HMI (see Figure 5), the most frequently stated emotion is found to have been "cool," while the least frequent was "amazing."

According to the statistics on positive attitudes toward YMI, the most frequent word expressing positive sentiment on the company's Facebook page was "best," and the least frequent was once again "amazing." As for SMI, positive sentiment was most frequently expressed as "best," and least frequently as "wonderful."

Positive emotion is quantified contextually, upon the lexical conversion of each sentence containing positive phrasing into a token of the positive wording. Although not all word-related content comprising the terms "love," "great," and "hate" was consistently positive or negative, the context is determined by the transition of tokenization on a survey-created and ranked exclusively phrase lexicon. For instance, in certain meaningful circumstances, likeability was quantified in terms of specific positive sentiment, whereas hate was quantified in terms of specific negative-sentiment content.

Figure. 5.
Data Intelligence for Positive Sentiment Toward HMI Company

In terms of negative emotion toward HMI (see Figure 6), the most frequently used term on the company's Facebook page is found to have been "don't like," and the least to have been "hate" and "horrible." As Figure 8 illustrates, the most frequently used negative expression was "dislike" in a context phrase.

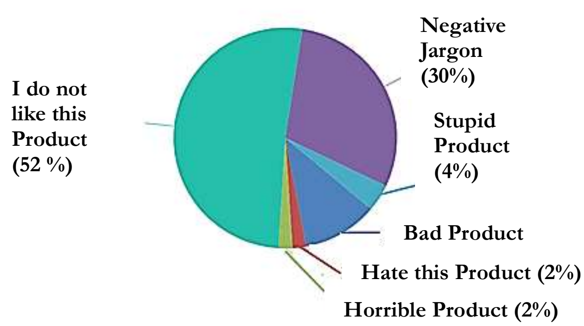

Figure. 6.
Data Intelligence for Negative Sentiment Toward HMI Company

The data for SMI indicates the most frequent expression of negative sentiment to have been "hate," with the term "stupid" being the least frequently used in a particular context. Conversely, for YMI, "stupid" was found to have been the most frequent expression of negative sentiment, while the least frequent one was the word "horrible."

Some form of word assimilation is often incorporated into a positive-content context. As a result, lexical comprehension must be severely restricted, such as the use of "but" or "by" in a sentence, which was considered in our transformation method.

Table 7 presents the utility of the amount of positive sentiment the conjoint analysis determined. This utility value designates the company's position from most to least positive, and the percentage conveys the regularity with which a positive level occurs. The greater the percentage, the more frequently a positive level has been encountered. Hence, a higher rate total indicates a more positive rate than a rival's. The benchmark assigns to HMI the highest rate value at 34.56, followed by SMI at 31.02 and YMI at 26.47 (as the lowest rate value).

Table 7.
Summary Of Data Analytics for Positive Sentiment

Positive Sentiment	Utility	HMI %	Rate	SMI %	Rate	YMI %	Rate
Love product	0,8907	26	20.8	18	14.4	19	15.2
Great product	-0,0496	4	0.16	3	0.12	3	0.12
Amazing product	-0,2846	1	0.2	8	1.6	1	0.2
Wonderful product	-0,5566	3	1.5	2	1	2	1
Like this product	0,4285	13	5.2	15	6	16	6.4
Nice product	-0,0831	15	1.2	10	0.8	10	0.8
Cool product	-0,1428	31	4.3	21	2.1	22	2.2
Best product	-0,2025	6	1.2	25	5	26	0.55
Total Value		34.56		31.02		26.47	

Table 8 displays the utility of the amount of negative sentiment according to conjoint analysis findings. The percentage indicates the frequency of occurrences in the context of a negative level. The larger the percentage, the greater the frequency of negative level contexts. A higher percentage indicates that one's rate is more negative than a rival's.

Table 8.
Summary of Data Analytics for Negative Sentiment

Negative Sentiment	Utility	HMI %	Rate	SMI %	Rate	YMI %	Rate
Hate this product	1,1692	2	2.2	44	48.4	26	28.6
Bad product	-0,4229	11	4.4	6	2.4	4	1.6
Don't like this product	-0,7463	52	36.4	29	20.3	17	11.9
Horrible product	-0,1194	2	0.2	1	0.1	1	0.1
Jargon/Sexual Object	-0,3085	30	9	17	5.1	10	3
Stupid product	0,4279	4	1.6	2	0.8	42	8.4
Total Value		53.8		77.1		53.6	

The benchmark assigns to SMI the highest negative rate value (77.1), followed by HMI (53.8) and YMI (53.6). As shown in Table 6, HMI is the most positive, which makes sense because HMI should be the least negative of the two case companies. It is important to underline that we gathered and processed the data on negativity and positivity independently.

From the results in Tables 7 and 8, we can infer that conjoint analysis is able to develop constructs for data analytics not only for determining sentiment positioning, but also for ranking such positivity or negativity. In other words, not only can we determine whether a brand is negatively or positively perceived, but also how negative or positive this perception is compared to the others evaluated (in this case, among competitors).

5. Conclusion

The outcome of conjoint analysis defining the utility of phrasing options is applicable to the first research question: "How does one assess a product, service, or brand using sentiment analysis and big data?" We present a method for determining which words are more psychologically meaningful, from most to least positive, based on statistical analysis gauging emotion. The fit model is depicted via the model statistics' results (Bak & Bartlomowicz, 2009). Table 7 and Table 8 indicate success in statistically quantifying the companies' positive and negative positions.

Table 6 assigns the most positive rate total (in the positive sentiment context), found via conjoint analysis, to HMI (34.56), followed by SMI (31.02) and YMI (26.47). This conclusion is statistically justified, since the utility of each positive phrase was quantified using a survey that eliminated questions such as which positive expression is most favorable to another positive expression. As for the negative sentiment context, Table 7 assigns the highest negative index to SMI (77.11), followed by HMI (53.8) and YMI (53.6).

In addressing the second research question, we gauged how preliminary data gathering could become a source of relevant market information. It follows that targeted data mining could enable the generation of more precise and focused context data. Also, data cleansing and data generation have been shown to immediately provide context for business-related data.

The process of implementing conjoint analysis provides a solution to the question regarding determining and ranking negative and positive terms based on frequency and intensity. By examining a company's positive and negative benchmarking positions, we can then examine the trends of a product, image, or service. In terms of businesses' decision-making prospects, this type of algorithm could help evaluate their performance with customers and competitive positions directly from users' perspective. We demonstrate that even with restricted resources, such as the amount of data available for retrieval, it is possible to gain essential knowledge.

As a direct commercial implication, our research manages to propose a thorough and replicable method quantifying and qualifying opinions expressed about a company's brand and products, i.e., the number of favorable and negative views expressed and how a company's performance compares to that of competitors. These types of insights are invaluable to businesses.

In using conjoint analysis as its analytical foundation, our study contributes to the body of research by providing an innovative viewpoint on data analysis. In contrast to conventional lexicon-based methodologies, particularly keywording techniques, not only does this study examine a sentence for its positivity or negativity, but it also describes how positive or negative it is. Even when combined, they can generally quantify the positive and negative aspects of a brand/product using the methodologies available.

Among this study's limitations is its inability to define the accuracy, volume growth, stability, and efficiency of data mining (Liu et al., 2018). It is critical to increase the number of comments on social media platforms. Future research should focus on evaluating data on a million-opinion mining scale accurately portraying the volume, velocity, and variety of big data (Sowmya et al., 2015), which this study did not accomplish.

Another limitation pertains to substitution of words. They were created using the wording dictionary on positive and negative sentiment we developed in response to a customer study about the value of terminology.

Customer perception within the wording construct should be measured in relation to service quality. For example, one could gauge how a sentence is related to the quality of products, processes, and services, as in "service science theory," which emphasizes the importance of evaluating the perception of service quality (Maglio & Spohrer, 2008). We did not directly measure the direct relationship between statements' negative and positive parts regarding service quality.

Lehtinen and Lehtinen (1991) found that aspects of service quality could take the form of quality in process, quality in output, and quality in environment. Future research should be able to measure the relationship between the construct of sentiment and service quality. For example, when someone expresses a negative sentiment, it could be determined, within the construction of service quality, whether the sentiment refers to service quality in the input, process, or environment, as suggested by Lehtinen and Lehtinen (1991).

Due to the complexity of language, the algorithm we used could benefit from numerous improvements in terms of comprehending the higher meaning of data polarity. Thus, future research should consider limiting the use of judgment rules in defining data-cleansing and data-formation decisions, by employing a more sophisticated formation algorithm capable of handling the complexity of language, which is typically found in complex-unit computing.

References

Akter, S., & Aziz, M. T. (2016). Sentiment analysis on facebook group using lexicon based approach. *3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*. doi:10.1109/ceeict.2016.7873080

Alsayat, A., & Elmitwally, N. (2020). A comprehensive study for Arabic sentiment analysis (Challenges and Applications). *Egyptian Informatics Journal*, 21(1), 7-12.

Alwakid, G., Osman, T., & Hughes-Roberts, T. (2017). Challenges in sentiment analysis for Arabic social networks. *Procedia Computer Science*, 117, 89-100.

Bak, A., & Bartlomowicz, T. (2009). Conjoint analysis method and its implementation in conjoint R package. Wroclaw: University of Economics.

Cattin, P., & Wittink, D. R. (1982). Commercial use of conjoint analysis: A survey. *Journal of marketing*, 46(3), 44-53.

Chiang, L. L. L., & Yang, C. S. (2018). Does country-of-origin brand personality generate retail customer lifetime value? A Big Data analytics approach. *Technological Forecasting and Social Change*, 130, 177-187.

Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research*, 55(17), 5142-5156.

Dong, R., O'Mahony, M. P., Schaal, M., McCarthy, K., & Smyth, B. (2015). Combining similarity and sentiment in opinion mining for product recommendation. *Journal of Intelligent Information Systems*, 46 (2), 285–312. doi:10.1007/s10844-015-0379-y

Eggers, F., Eggers, F., & Kraus, S. (2016). Entrepreneurial branding: measuring consumer preferences through choice-based conjoint analysis. *International Entrepreneurship and Management Journal*, 12(2), 427-444.

Green, P. E., & Krieger, A. M. (1993). Conjoint analysis with product-positioning applications. *Handbooks in operations research and management science*, 5, 467-515.

Green, P. E., & Krieger, A. M. (2002). What's right with conjoint analysis?. *Marketing Research*, 14(1), 24.

Hauser, J. R., & Rao, V. R. (2004). Conjoint Analysis, Related Modeling, and Applications. *Marketing Research and Modeling: Progress and Prospects*, 141–168. doi:10.1007/978-0-387-28692-1_7

Helmi, R. L. (2020). Knowledge Management Enabler (KME) to Promote Innovation Capabilities in Public R&D Centers in Indonesia. *The Asian Journal of Technology Management,* 13(2), 98-112.

Hemalatha, I., Varma, G. S., & Govardhan, A. (2013). Sentiment classification in online reviews using FRN algorithm. 357-362.

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences*, 30(4), 330-338.

Jian Jin, & Ping Ji. (2015). Mining online product reviews to identify consumers' fine-grained concerns. *12th International Symposium on Operations Research and Its Applications in Engineering, Technology and Management (ISORA 2015).*

doi:10.1049/cp.2015.0622

Kencanasari, R. A. M., Dhewanto, W., & Rustiadi, S. (2021). Digital Product Perception and User Satisfaction Relationship: Can They Create Feedback Intention?. *The Asian Journal of Technology Management*, 14(2), 109-127.

Khedr, A. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, 9(7), 22.

Khoo, C. S. G., Nourbakhsh, A., & Na, J. C. (2012). Sentiment analysis of online news text: a case study of appraisal theory. *Online Information Review*, 36(6), 858-878.

Kralj Novak, P., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PloS one*, 10(12), e0144296.

Kumar, S., & Kaur, K. (2016). Review of Data mining (Knowledge discovery) in the Future. *International Journal of Advanced Research in Computer Science*, 7(6).

Lehtinen, U., & Lehtinen, J. R. (1991). Two approaches to service quality dimensions. *Service Industries Journal*, 11(3), 287-303.

Lee, E., Lee, J. A., Moon, J. H., & Sung, Y. (2015). Pictures speak louder than words: Motivations for using Instagram. *Cyberpsychology, behavior, and social networking*, 18(9), 552-556.

Lim, S. Y., Harun, U. B., Gobil, A. R., Mustafa, N. A., Zahid, N. A., Amin-Nordin, S., ... & Shohaimi, S. (2021). Measuring customer satisfaction on the cleanliness of food premises using fuzzy conjoint analysis: A pilot test. *Plos one*, 16(9), e0256896.

Lin, D., Cao, D., Lv, Y., & Cai, Z. (2017). GIF video sentiment detection using semantic sequence. *Mathematical Problems in Engineering*, 2017, 1-11.

Liu, X., Zhou, Y., & Chen, X. (2018). Mining outlier data in mobile internet-based large real-time databases. Complexity, 2018, 1-12

Maglio, P. P., & Spohrer, J. (2008). Fundamentals of service science. *Journal of the academy of marketing science*, 36(1), 18-20.

Mariani, M. M., & Wamba, S. F. (2020). Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies. *Journal of Business Research*, 121, 338-352.

Marshall, D., Bridges, J. F., Hauber, B., Cameron, R., Donnalley, L., Fyie, K., & Reed Johnson, F. (2010). Conjoint analysis applications in health—how are studies being designed and reported?. The Patient: *Patient-Centered Outcomes Research*, 3(4), 249-256.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.

Papadima, G., Genitsaris, E., Karagiotas, I., Naniopoulos, A., & Nalmpantis, D. (2020). Investigation of acceptance of driverless buses in the city of Trikala and optimization of the service using Conjoint Analysis. *Utilities Policy*, 62, 100994.

Reagan, A. J., Danforth, C. M., Tivnan, B., Williams, J. R., & Dodds, P. S. (2017). Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. *EPJ Data Science*, 6, 1-21.

Sandoval-Almazan, R., & Valle-Cruz, D. (2018). Facebook impact and sentiment analysis on political campaigns. *In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 1-7.

Saura, J. R., Herráez, B. R., & Reyes-Menendez, A. (2019). Comparing a traditional approach for financial Brand Communication Analysis with a Big Data Analytics technique. *IEEE Access*, 7, 37100-37108.

Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Wai, P. S., Chung, Y. W., ... & Al-Garadi, M. A. (2018). Sentiment analysis of big data: Methods, applications, and open challenges. *IEEE Access*, 6, 37807-37827.

Shirdastian, H., Laroche, M., & Richard, M. O. (2019). Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter. *International Journal of Information Management*, 48, 291-307.

Smetanin, S. (2020). The applications of sentiment analysis for Russian language texts: Current challenges and future perspectives. *IEEE Access*, 8, 110693-110719.

Soliman, T. H., Elmasry, M. A., Hedar, A., & Doss, M. M. (2014). Sentiment analysis of Arabic slang comments on facebook. *International Journal of Computers & Technology*, 12(5), 3470-3478.

Sowmya, Y., Ratna, M. N., & Bindu, C. S. (2015). A Review on Big Data Mining, Distributed Programming Frameworks and Privacy Preserving Data Mining Techniques. *International journal of advanced research in computer science*, 6(1), 121-126.

Sun, X., Wagner, A. L., Ji, J., Huang, Z., Zikmund-Fisher, B. J., Boulton, M. L., ... & Prosser, L. A. (2020). A conjoint analysis of stated vaccine preferences in Shanghai, China. *Vaccine*, 38(6), 1520-1525.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of marketing research*, 51(4), 463-479.

Tripathi, S. N., & Siddiqui, M. H. (2010). An empirical study of tourist preferences using conjoint analysis. *International Journal of Business Science & Applied Management (IJBSAM)*, 5(2), 1-16.

Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2013). Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. *In IISA 2013*, 1-6.

Wang, L., & Wang, G. (2015). Data mining applications in big data. *Computer Engineering and Applications Journal*, 4(3), 143-152.

Youn, D. M., Lee, Y. H., & Lee, B. G. (2020). Proposal of Brand Evaluation Map through Big Data: Focus on The Hyundai Motor's Product Evaluation. *Journal of Information Technology Services*, 19(4), 1-11.